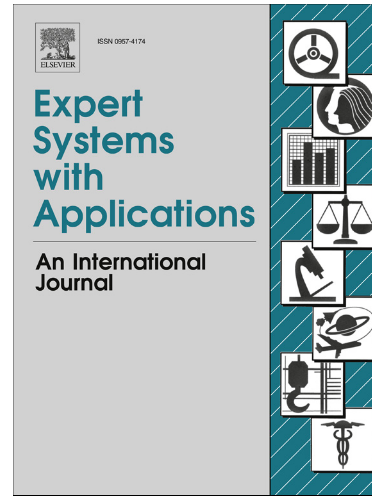
Accepted Manuscript



Unsupervised Feature Selection using an Improved version of Differential Evolution

Tapas Bhadra, Sanghamitra Bandyopadhyay

PII:	S0957-4174(14)00783-0
DOI:	http://dx.doi.org/10.1016/j.eswa.2014.12.010
Reference:	ESWA 9733
To appear in:	*Expert Systems with Applications*

Please cite this article as: Bhadra, T., Bandyopadhyay, S., Unsupervised Feature Selection using an Improved version of Differential Evolution, *Expert Systems with Applications* (2014), doi: http://dx.doi.org/10.1016/j.eswa.2014.12.010

Unsupervised Feature Selection using an Improved version of Differential Evolution

Tapas Bhadra*, Sanghamitra Bandyopadhyay

Machine Intelligence Unit, Indian Statistical Institute, Kolkata-700108, India.

Abstract

In this article, an unsupervised feature selection algorithm is proposed using an improved version of a recently developed Differential Evolution technique called MoDE. The proposed algorithm produces an optimal feature subset while optimizing three criteria, namely, the average standard deviation of the selected feature subset, the average dissimilarity of the selected features, and the average similarity of non-selected features with respect to their first nearest neighbor selected features. Normalized mutual information score is employed for computing both the similarity as well as the dissimilarity measures. The experimental results confirm the superiority of the proposed algorithm over the other state-of-the-art unsupervised feature selection algorithms for eight different kinds of datasets with the number of points ranging from 80 to 6238 and the number of dimensions ranging from 30 to 649.

Keywords: Pattern recognition; Unsupervised feature selection; Mutual information; Normalized mutual information; Differential evolution.

1. Introduction

Pattern recognition techniques are widely adopted to map a real-life problem into an equivalent machine learning one (Duda et al., 2001). Data collection, feature analysis, and classification, are the three commonly used steps of any standard pattern recognition system (Pal & Mitra, 2004). In the era of large datasets, it is a natural phenomenon that some features are often redundant or correlated with other features. In addition to these, some noisy features may be introduced at the time of data collection. Thus, discarding such redundant, noisy features is necessary for enhancing the prominent characteristics of the data. As compared to the total feature space, a smaller number of features not only improves the computation speed up, but also provides a more compact model with better generalization capability. Dimensionality can be reduced in two ways, namely, feature selection and feature extraction. Depending upon the learning principle, feature selection can be of two types, supervised and unsupervised feature selection (Dash & Liu, 1997). Supervised feature selection is employed when the class information leads the path for choosing the reduced feature set; otherwise one uses unsupervised learning approach. Based on the feature assessment method, feature selection algorithms can also be classified into two groups, namely, filter and wrapper methods. In the former approach, the potential power of each candidate feature subset is assessed using some statistical measures that include chi-square test (Jin et al., 2006), Wilcoxon Mann–Whitney test (Liao et al., 2007), t-test (Hua et al., 2008), mutual information (Battiti, 1994; Kwak & Choi, 2002; Peng et al., 2005), normalized mutual information (Estévez

*Corresponding author at: Machine Intelligence Unit, Indian Statistical Institute, Kolkata- 700108, India, Tel.: +91-33-2575-3114; Fax.: +91-33-2578-3357.

Email addresses: tapas.bhadra@isical.ac.in (Tapas Bhadra), sanghami@isical.ac.in (Sanghamitra Bandyopadhyay)